

Customer Lifetime Value Prediction

Project Report

Introduction

This project aimed to predict Customer Lifetime Value (CLTV) based on historical transactional data. CLTV is a crucial metric for businesses, aiding in identifying valuable customers and optimizing marketing strategies.

Abstract

This report details the process of building a Customer Lifetime Value (CLTV) prediction model using the Online Retail II dataset. It covers data preprocessing, feature engineering using RFM analysis, and model training with a RandomForestRegressor. The project demonstrates an end-to-end approach to CLTV prediction, providing insights for customer segmentation and targeted marketing.

Tools Used

- Python (Sklearn, XGBoost)
- Excel

Steps Involved in Building the Project

Data Preprocessing

1. **Data Loading and Combination:** Combined data from two Excel sheets (Year 2009-2010 and Year 2010-2011).
2. **Handling Missing Values:** Dropped rows with missing Customer ID.

3. **Date Conversion:** Converted `InvoiceDate` to datetime objects.

4. **Total Price Calculation:** Calculated `TotalPrice` as `Quantity * Price`.

Feature Engineering (RFM Analysis)

Recency, Frequency, and Monetary (RFM) features were engineered for each customer:

* **Recency:** Days since last purchase. * **Frequency:** Total unique purchases. *

Monetary: Total amount spent. * **Average Order Value (AOV):** Monetary / Frequency.

Model Training and Evaluation

A `RandomForestRegressor` model was trained to predict CLTV (defined as Monetary value for this demonstration). The model was evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

Conclusion

This project successfully built a basic CLTV prediction model, demonstrating data acquisition, preprocessing, feature engineering, and model training. The generated outputs can be used for customer segmentation and targeted marketing efforts.

Results

Evaluation Metrics: - Mean Absolute Error (MAE): 440.36 - Root Mean Squared Error (RMSE): 6775.13

Deliverables

- `Python notebook`: (`Python_notebook.py`)
- `Trained model + visualizations`: (`cltv_model.joblib`, `rfm_distributions.png`, `actual_vs_predicted_cltv.png`)
- `Final LTV prediction CSV`: (`Final_LTV_prediction.csv`)